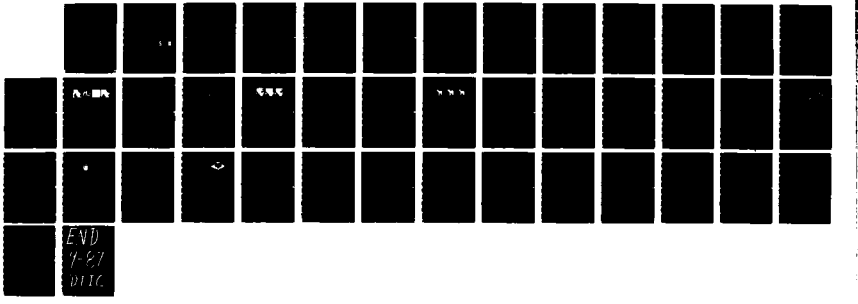
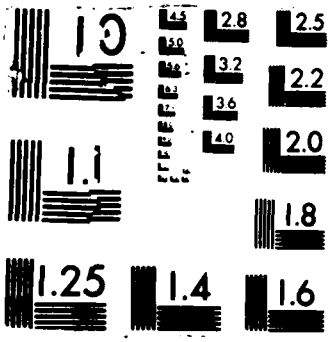
AD-A183 807 PROBABILISTIC SOLUTION OF ILL-POSED PROBLEMS IN COMPUTATIONAL VISION(U) MASSACHUSETTS INST OF TECH CAMBRIDGE ARTIFICIAL INTELLIGENCE LAB 1/1

UNCLASSIFIED J MARROQUIN ET AL MAR 87 AI-M-897 F/G 12/9 NL



END
7-87
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0

(12)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1 REPORT NUMBER	2 GOVT ACCESSION NO	3 RECIPIENT'S CATALOG NUMBER
897		

4 TITLE (and Subtitle)

Probabilistic Solution of Ill-Posed Problems in Computational Vision

5 TYPE OF REPORT & PERIOD COVERED

Memo

6 PERFORMING ORG. REPORT NUMBER

7 AUTHOR(s)

J. Marroquin
S. Mitter
T. Poggio

8 CONTRACT OR GRANT NUMBER(s)

N00014-85-K-0124
ARO-DAAG29-84-K-0005
NR SRO-202

9 PERFORMING ORGANIZATION NAME AND ADDRESS

Artificial Intelligence Laboratory
545 Technology Square
Cambridge, MA 02139

10 PROGRAM ELEMENT PROJECT, TASK AREA & WORK UNIT NUMBERS

11 CONTROLLING OFFICE NAME AND ADDRESS

Advanced Research Projects Agency
1400 Wilson Blvd.
Arlington, VA 22209

12 REPORT DATE

March 1987

13 NUMBER OF PAGES

40

14 MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)

Office of Naval Research
Information Systems
Arlington, VA 22217

15 SECURITY CLASS. (of this report)

UNCLASSIFIED

15a DECLASSIFICATION/DOWNGRADING SCHEDULE

16 DISTRIBUTION STATEMENT (of this Report)

Distribution is unlimited.

17 DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18 SUPPLEMENTARY NOTES

None

19 KEY WORDS (Continue on reverse side if necessary and identify by block number)

Ill-posed problems, Regularization, Stochastic methods
Computer vision, SIMD architecture

20 ABSTRACT (Continue on reverse side if necessary and identify by block number)

Computational vision is a set of inverse problems. We review standard regularization theory, discuss its limitations, and present new stochastic (in particular, Bayesian) methods for their solution. We derive efficient algorithms and describe parallel implementations on digital parallel SIMD architectures, as well as a new class of parallel hybrid computers.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A. I. Memo 897 March 1987

Probabilistic Solution of Ill–Posed Problems in Computational Vision

J. Marroquin[3], S. Mitter[2], and T. Poggio[1]

Abstract: We formulate several problems in early vision as inverse problems. Among the solution methods we review standard regularization theory, discuss its limitations, and present new stochastic (in particular, Bayesian) techniques based on Markov Random Field models for their solution. We derive efficient algorithms and describe parallel implementations on digital parallel SIMD architectures, as well as a new class of parallel hybrid computers that mix digital with analog components.

Acknowledgements. This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the Laboratory's artificial intelligence research is provided in part by the Advanced Research Products Agency of the Department of Defense under Office of Naval Research contract N00014-85-K-0124. Some support for Tomaso Poggio is provided by a gift from the Artificial Intelligence Division of the Hughes Aircraft Corporation. A version of this memo was published in the Proceedings of the Image Understanding Workshop held in Miami, 1985. Support for J. Marroquin and S. Mitter was supplied by the Army Research Office under contract ARO-DAAG29-84-K-0005. This research was also supported in part by Office of Naval Research contract NR SRO-202.

[1]Artificial Intelligence Laboratory. [2]Laboratory for Information and Decision Systems. [3]Centro de Investigacion en Matematicas; Apdo.Postal 402; Guanajuato, Gto; Mexico.

1. Introduction

1.1. Computational Vision

Computational vision denotes a new field in artificial intelligence that has developed in the last 15 years. Its two main goals are to develop image understanding systems which automatically provide scene descriptions from real images, and to understand biological vision. Its main focus is on theoretical studies of vision, considered as an information processing task.

Since at least the work of David Marr (Marr, 1982; see also Marr and Poggio, 1977), it has been customary to consider vision as an information processing system that can be divided into several modules at different theoretical levels, at least as a first approximation. In particular, Marr suggested that the goal of the first step of vision is to obtain descriptions of physical properties of three-dimensional surfaces around the viewer such as distance, orientation, texture, and reflectance. This first step of vision, up to what has been called *2-1/2 D sketch* or *intrinsic images*, is mainly bottom-up, relying on general knowledge but no special high-level information about the scene to be analyzed.

The first part of vision–from images to surfaces–has been termed *early vision*. Although this point-of-view has been embraced widely (see a set of recent reviews, e.g., Brown, 1984; Brady, 1981; Barrow and Tannenbaum, 1981; Poggio, 1984), it is important to observe that its correctness is still to be proven. In particular, it is still unclear what the nature of the 2-1/2-D sketch representation is, how different visual modules interact, how their output is fused and what is the role of high-level knowledge on early visual processes. The critical problem of the organization of vision and of the control of the flow of information from the different modules and how high-level knowledge is used is still very much an open problem.

In this paper, we do not consider this larger issue. Our point-of-view is that a rigorous analysis of individual modules of vision is bound to play an important role in any full theory of vision.

1.2. Early Vision

Early vision consists of a set of processes that recover physical properties of visible three-dimensional surfaces from the two-dimensional images. Computational, biological and epistemological arguments (see Marr and Poggio, 1977)

- Edge Detection
- Spatio-temporal interpolation and approximation
- Computation of optical flow
- Computation of lightness and albedo
- Shape from contours
- Shape from texture
- Shape from shading
- Binocular Stereo
- Structure from motion
- Structure from stereo
- Surface reconstruction
- Computation of Surface Color

Table 1. Examples of early vision processes.

suggest that early vision processes are generic ones that correspond to conceptually independent modules that can be studied, at least to a very first approximation, in isolation. Table 1 shows a list of some of the early vision modules.

The standard definition of computational vision is that it is inverse optics. The direct problem–the problem of classical optics–or computer graphics–is to determine the images of three-dimensional objects. Computational vision is confronted with inverse problems of recovering surfaces from images. Much information is lost during the imaging process that projects a three-dimensional world into two-dimensional arrays (images). As a consequence, vision must rely on natural constraints, that is, general assumptions about the physical world to derive an unambiguous output. This is typical of many inverse problems in mathematics and physics.

In fact, the common characteristics of most early vision problems, in a sense their deep structure, can be formalized: *early vision problems are ill-posed in the sense defined by Hadamard.* A problem is well-posed when its solution (a) exists, (b) is unique and (c) depends continuously on the initial data. Ill-posed problems fail to satisfy one or more of these criteria.

Bertero, Poggio and Torre (1987) show precisely the mathematically ill-posed structure of several problems listed in Table 1 (see also Poggio and Torre,

1984.) The recognition that early vision problems are ill-posed suggests immediately the use of regularization methods developed in mathematics and mathematical physics for solving the ill-posed problems of early vision (Poggio and Torre, 1984). Without an explicit connection with regularization techniques, B. Horn (1986) had earlier approached several problems in vision from a very similar point of view, using minimization techniques for their solution.

1.3. Standard Regularization in Early Vision

The main idea for "solving" ill-posed problems is to restrict the class of admissible solutions by introducing suitable *a priori* knowledge. In standard regularization methods, due mainly to Tikhonov, the regularization of the ill-posed problem of finding z from the data y :

$$Az = y \tag{1}$$

results in finding z that minimizes

$$\|Az - y\|^2 + \lambda \|Pz\|^2, \tag{2}$$

where λ is a so-called regularization parameter, $\|\cdot\|$ is the norm and $\|Pz\|$ is called a stabilizing functional. In standard regularization theory, A is a linear operator, the norms are quadratic and P is linear. In this method, λ controls the compromise between the degree of regularization of a solution and its closeness to the data (the first term in equation 2). P embeds the physical constraints of the problem. It can be shown for quadratic variational principles that under mild conditions the solution space is convex and a unique solution exists.

Poggio et al (1984, 1985) show that several problems in early vision can be "solved" by standard regularization techniques and Bertero et al. (1987) give a rigorous analysis. Surface reconstruction, optical flow at each point in the image, optical flow along contours, color, and stereo can be computed by using standard regularization techniques. Variational principles that are not exactly quadratic but have the same form as equation 2 can be used for other problems in early vision. The main results of Tikhonov can, in fact, be extended to some cases in which the operators A and P are nonlinear, provided they satisfy certain conditions (Morozov, 1984.)

Standard regularization methods can be implemented very efficiently by parallel architectures of the fine-grain type, such as the Connection Machine (Hillis, 1985). Minimizing formula (2) generates a convolution operator when certain conditions are met (Poggio et al., 1986). Analog networks, either electrical or chemical, can also be a natural way of solving the variational principles dictated by standard regularization theory (Poggio and Koch, 1984, 1985). A list of the problems that can be regularized by standard regularization theory or

Problem	Regularization Principle
Edge detection	$\int \left[(Sf - i)^2 + \lambda (f_{xx})^2 \right] dx$
Optical Flow (area based)	$\int \left[(i_x u + i_y v + i_t)^2 + \lambda (u_x^2 + u_y^2 + v_x^2 + v_y^2) \right] dx dy$
Optical Flow (contour based)	$\int \left[(\mathbf{V} \cdot \mathbf{N} - v^N)^2 + \lambda (\frac{\partial}{\partial s} \mathbf{V})^2 \right] ds$
Surface Reconstruction	$\int \left[(S \cdot f - d)^2 + \lambda (f_{xx}^2 + 2 f_{xy}^2 + f_{yy}^2)^2 \right] dx\ dy$
Spatiotemporal approximation	$\int \left[(\mathbf{S} f - i)^2 + \lambda (\nabla f \cdot \mathbf{V} + f_t)^2 \right] dx\ dy\ dt$
Color	$\|I^\nu - Az\|^2 + \lambda \|Pz\|^2$
Shape from Shading	$\int \left[(E - R(f,g))^2 + \lambda (f_x{}^2 + f_y{}^2 + g_x{}^2 + g_y{}^2) \right] dx\ dy$
Stereo	$\int \left\{ \left[\nabla^2 G * \Big(L(x,y) - R(x + d(x,y), y) \Big) \right]^2 + \lambda (\nabla d)^2 \right\} dx\ dy$

Table 2. Regularization in early vision.

slightly non-linear versions of it are listed in Table 2, together with the associated regularization principle.

1.4. Limitations of Standard Regularization Theory

This new theoretical framework for early vision shows clearly not only the attractions, but also the limitations, that are intrinsic to the standard Tikhonov form of regularization theory. Standard regularization methods lead to satisfactory solutions of early vision problems but cannot deal effectively and directly with a few general problems such as *discontinuities* and *fusion of information from multiple modules*.

Standard regularization theory with linear A and P is equivalent to restricting the space of solution to generalized splines, whose order depends on the order of the stabilizer P. This means that in some cases the solution is too

smooth, and cannot be faithful in locations where discontinuities are present. In optical flow, surface reconstruction and stereo, discontinuities are in fact not only present, but also the most critical locations for subsequent visual information processing. Standard regularization cannot deal well with another critical problem of vision, the problem of fusing information from different early vision modules. Since the regularizing principles of the standard theory are quadratic, they lead to linear Euler-Lagrange equations. The output of different modules can therefore be combined only in a linear way. Terzopoulous (1984; see also Poggio et al., 1985) has shown how standard regularization techniques can be used in the presence of discontinuities in the case of surface interpolation. After standard regularization, locations where the solution f originates a large error in the second term of equation 2, are identified (this needs setting a threshold for the error in smoothness). A second regularization step is then performed using the location of discontinuities as boundary conditions.

A similar method could be used for fusing information from multiple sources: a regularizing step could be performed and locations where terms of the type of the first term of equation 2 give large errors would be identified. A decision step would then follow by setting appropriately various controlling parameters in those locations, therefore weighting in an appropriate way (for instance, vetoing some of) the various contributing processes.

One would like, however, a more comprehensive and coherent theory capable of dealing directly with the problem of discontinuities and the problem of fusing information. So the challenge for a regularization theory of early vision is to extend it beyond standard regularization methods and their most obvious non-linear versions.

1.5. Stochastic Route to Regularizing Early Vision

In this paper, we will outline a rigorous approach to overcome part of the ill-posedness of vision problems, based on Bayes estimation and Markov Random Field models, that effectively deals with the problems faced by the standard regularization approach. In this approach, the *a priori* knowledge is represented in terms of an appropriate probability distribution, whereas in standard regularization *a priori* knowledge leads to restrictions on the solution space. This distribution, together with a probabilistic description of the noise that corrupts the observations, allows one to use Bayes theory to compute the posterior distribution $P_{f|g}$, which represents the likelihood of a solution f given the observations g. In this way, we can solve the reconstruction problem by finding the estimate $\hat{f}$ which either maximizes this a posteriori probability distribution (the so called

Maximum a Posteriori or MAP estimate), or minimizes the expected value (with respect to $P_{f|g}$) of an appropriate error function. The class of solutions that can be obtained in this way is much larger than in standard regularization. In particular, we will show under which conditions this new method leads to solutions that are of the standard regularization type (see Section 3).

The price to be paid for this increased flexibility is computational complexity. New parallel architectures and possibly hybrid computers of the digital-analog type promise however to deal effectively with the computational requirements of the methods proposed here. We will discuss these new parallel architectures in some detail at the end of the paper.

2. Probabilistic Models

In the rest of the paper we are concetrating on one specific problem in early vision, the problem of surface reconstruction.

The key to the success in the use of the probabilistic approach is our ability to find a class of stochastic models (that is, random fields) that have the following characteristics:

The probabilistic dependencies between the elements of the field should be local. This condition is necessary if the field is to be used to model surfaces that are only piecewise smooth; besides, if it is satisfied, the reconstruction algorithms are likely to be distributed, and thus, efficiently implementable in parallel hardware.

The class should be rich enough, so that a wide variety of qualitatively different behaviors can be modeled.

The relation between the parameters of the models and the characteristics of the corresponding sample fields should be relatively transparent, so that the models are easy to specify.

It should be possible to represent the prior probability distribution P_f explicitly, so that Bayes theory can be applied.

It should be possible to specify efficient Monte Carlo procedures, both for generating sample fields from the distribution, so that the capability of the model to represent our prior knowledge can be verified, and to compute the optimal estimators.

A class of random fields that satisfies these requirements is the class of Markov Random Fields (MRF's) on finite lattices (see Wong, 1968 and Woods, 1972). A MRF has the property that the probability distribution of the configurations of the field can always be expressed in the form of a Gibbs distribution:

$$P_f(f) = \frac{1}{Z} e^{-\frac{1}{T_0} U(f)} \tag{3}$$

where Z is a normalizing constant, T_0 is a parameter (known as the "natural temperature" of the field) and the "Energy function" $U(f)$ is of the form:

$$U(f) = \sum_C V_C(f) \tag{4}$$

where C ranges over the "cliques" associated with the neighborhood system of the field, and the potentials $V_C(f)$ are functions supported on them (a clique is either a single site, or a set of sites such that any two sites belonging to it are neighbors of each other).

As an example, the behavior of piecewise constant functions can be modeled using first order MRF models on a finite lattice L with generalized Ising potentials (Geman and Geman, 1984):

$$V_C(f_i, f_j) = \begin{cases} -1, & \text{if } |i-j| = 1 \text{ and } f_i = f_j \\ 1, & \text{if } |i-j| = 1 \text{ and } f_i \neq f_j \\ 0, & \text{otherwise.} \end{cases}$$

$$f_i \in Q_i = \{q_1, \ldots, q_M\} \quad \text{for all } i \in L \tag{5}$$

We will use a free boundary model, so that the neighborhood size for a given site will be: 4, if it is in the interior of the lattice; 3, if it lies at a boundary, but not at a corner, and 2 for the corners.

The Gibbs distribution:

$$P_f(f) = \frac{1}{Z} \exp\left[-\frac{1}{T_0} U_0(f)\right]$$

$$U_0(f) = \sum_{i,j} V(f_i, f_j) \tag{6}$$

defines a one parameter family of models (indexed by T_0) describing piecewise constant patterns with varying degrees of granularity.

We will assume that the available observations g are obtained from a typical realization f of the field by a degrading operation (such as sampling) followed by corruption with i.i.d. noise (the form of whose distribution is known), so that the conditional distribution can be written as:

$$P_{g|f}(g; f) = \exp(-\alpha \hat{\epsilon}) \tag{7}$$

with $\hat{\epsilon} = \sum_{i \in S} \Phi_i(f, g_i)$, where $\{\Phi_i\}$ are some known functions, and α is a parameter.

The posterior distribution is obtained from Bayes rule:

$$P_{f|g}(f;g) = \frac{1}{Z_P} \exp\left[-U_P(f;g)\right] \tag{8}$$

with

$$U_P(f;g) = \frac{1}{T_0} U_0(f) + \sum_{i \in S} \Phi(f, g_i) \tag{9}$$

For example, in the case of binary fields ($M = 2$) with the observations taken as the output of a binary symmetric channel (BSC) with error rate ϵ (Gallager, 1975), we have:

$$P(g_i|f_i) = \begin{cases} (1-\epsilon), & \text{for } g_i = f_i \\ \epsilon, & \text{for } g_i \neq f_i \end{cases} \tag{10}$$

The posterior energy reduces to:

$$U_P(f;g) = \frac{1}{T_0} \sum_{i,j} V(f_i, f_j) + \alpha \sum_i (1 - \delta(f_i - g_i)) \tag{11}$$

where $f_i \in \{q_1, q_2\}$;

$$\delta(a) = \begin{cases} 1, & \text{if } a = 0 \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

and

$$\alpha = \ln\left(\frac{1-\epsilon}{\epsilon}\right). \tag{12}$$

3. Cost Functionals

The Bayesian approach to the solution of reconstruction problems has been adopted by several researchers. In most cases, the criterion for selecting the optimal estimate has been the maximization of the posterior probability (the Maximum a Posteriori or MAP estimate). It has been used, for example, by Geman and Geman (1984) for the restoration of piecewise constant images; by Grenander (1984) for pattern reconstruction, and by Elliot et. al. (1983) and Hansen and Elliot (1982) for the segmentation of textured images (a similar criterion — the maximization of a suitably defined likelihood function — has been used by Cohen and Cooper (1984) for the same purposes).

In some other cases, a performance criterion, such as the minimization of the mean squared error has been implicitly used for the estimation of particular classes of fields. For example, for continuous-valued fields with exponential autocorrelation functions, corrupted by additive white Gaussian noise, Nahi and

Assefi (1972) and Habibi (1972) have used causal linear models and optimal (Kalman) linear filters for solving the reconstruction problem.

The minimization of the expected value of error functionals, however, has not been used as an explicit criterion for designing optimal estimators in the general case. We will show that this design criterion is in fact more appropriate in our case, for the following reasons:

It permits one to adapt the estimator to each particular problem.

It is in closer agreement with one's intuitive assessment of the performance of an estimator.

It leads to attractive computational schemes.

As an example, we will now propose design criteria for two particular problems: image segmentation and surface reconstruction.

Consider a field f with N elements each of which can belong to one of a finite set Q_i of classes. Let f_i denote the class to which the i^{th} element belongs. The segmentation problem is to estimate f from a set of observations $\{g_1, \ldots, g_p\}$. Note that f_i does not necessarily correspond to the image intensity. It may represent, for example, the texture class for a region in the image (as in Elliot et. al., 1983), etc.

A reasonable criterion for the performance of an estimate $\hat{f}$ is the number of elements that are not classified correctly. Therefore, we define the segmentation error e_s as:

$$e_s(f, \hat{f}) = \sum_{i=1}^{N} (1 - \delta(f_i - \hat{f}_i)), \qquad f_i, \hat{f}_i \in Q_i. \tag{14}$$

In the case of the reconstruction problem, an estimate $\hat{f}$ should be considered "good" if it is close to f in the ordinary sense, so that the total squared error

$$e_r(f, \hat{f}) = \sum_{i=1}^{N} (f_i - \hat{f}_i)^2 \tag{15}$$

will be a reasonable measure for its performance.

To derive the optimal estimators with respect to the criteria stated above, we first present the general result (which can be found, for example in Abend, 1968) which states that if the posterior marginal distributions for every element of the field are known, the optimal Bayesian estimator with respect to any additive, positive definite cost functional C may be found by independently minimizing the marginal expected cost for each element.

In more precise terms, we will consider cost functionals $C(f, \hat{f})$ of the form:

$$C(f, \hat{f}) = \sum_{i \in L} C_i(f_i, \hat{f}_i) \tag{16}$$

with

$$C_i(a, b) \begin{cases} = 0 & \text{if } a = b \\ > 0 & \text{if } a \neq b, \text{ for all } i. \end{cases} \tag{17}$$

We will assume that the value of each element f_i of the field f is constrained to belong to some finite set Q_i (the generalization to the case of compact sets is straightforward). The Optimal Bayesian estimator $\hat{f}^*$ with respect to the cost functional C is defined as the global minimizer of the expected value of C over all possible f and g. One can prove that this estimate can be found by minimizing independently the marginal expected cost for each element, i.e.,

$$\hat{f}_i^* = q \quad : \quad \sum_{r \in Q_i} C_i(r, q) P_i(r|g) \leq \sum_{r \in Q_i} C_i(r, s) P_i(r|g)$$
$$\text{for all } s \neq q, \text{ and for all } i \in L \tag{18}$$

where $P_i(r|g)$ is the posterior marginal distribution of the element i:

$$P_i(r|g) = \sum_{f : f_i = r} P_{f|g}(f; g) \tag{19}$$

The optimal estimators for the error criteria defined above can be easily derived from this result:

In the case of the segmentation problem, we get that

$$\hat{f}_i^* = q \in Q_i \quad : \quad P_i(q|g) \geq P_i(s|g)$$
$$\text{for all } s \neq q. \tag{20}$$

We will call this estimate the "Maximizer of the Posterior Marginals" ($\hat{f}_{MPM}$).

For the reconstruction problem, the optimal estimate is:

$$f_i^* = q \in Q_i \quad : \quad (f_i - q)^2 \leq (f_i - s)^2$$
$$\text{for all } s \neq q \tag{21}$$

We will call this estimate the "Thresholded Posterior Mean" ($\hat{f}_{TPM}$).

The main obstacle for the practical application of these results lies in the formidable computational cost associated with the exact computation of the marginals and the mean of the posterior distribution given by (9), even for lattices of moderate size. In the next section we will present a general distributed procedure that will permit us to approximate these quantities as precisely as we may want.

4. Algorithms

The algorithms that we will propose are based on the use of the Metropolis (Metropolis et al., 1956) or Gibbs Sampler (Geman and Geman, 1984) schemes, to simulate the equilibrium behavior of the coupled MRF described by Equation (9). We recall that the Markov chain generated by these algorithms is regular, and their invariant measure is the posterior distribution $P_{f|g}$. The law of large numbers for regular chains (see, for example, Kemeny and Snell, 1960) establishes that the fraction of time that the chain will spend on a given state f will tend to $P_{f|g}(f;g)$ as the number of steps gets large, independently of the initial state. This means that we can approximate the posterior marginals by:

$$P_i(q|g) \approx \frac{1}{k-n} \sum_{t=k}^{n} \delta(f_i^{(t)} - q) \tag{22}$$

and $\bar{f}$ by:

$$\bar{f}_i \approx \frac{1}{n-k} \sum_{t=k}^{n} f_i^{(t)} \tag{23}$$

where $f^{(t)}$ is the configuration generated by the Metropolis algorithm at time t, and k is the time required for the system to be in thermal equilibrium. From these values, $\hat{f}_{MPM}$ and $\hat{f}_{TPM}$ can be easily computed using (20) and (21).

This procedure is related to the use of simulated annealing for finding the global minimum of U_P (i.e., the MAP estimate: see Geman and Geman, 1984). In our case, however, we are interested in gathering statistics about the equilibrium behavior of the coupled field *at a fixed temperature* $T = 1$, rather than in finding the ground state of the system. This fact gives our procedure some distinct advantages:

1. It is difficult to determine in general the descent rate of the temperature (annealing schedule) that will guarantee the convergence of the annealing process in a reasonable time (it usually involves a trial and error procedure; though Geman and Geman (1984) gave theoretical bounds for the annealing schedule, these bounds are impractical). Since we are running the Metropolis algorithm at a fixed temperature, this issue becomes irrelevant.

2. Since in our case we are using a Monte Carlo procedure to approximate the values of some integrals, we should expect a nice convergence behavior, in the sense that coarse approximations can be computed very rapidly, and then refined to an arbitrary precision (in fact, it can be proved (see Feller, 1950) that the expected value of the squared error of the estimates (22) and (23) is inversely proportional to n).

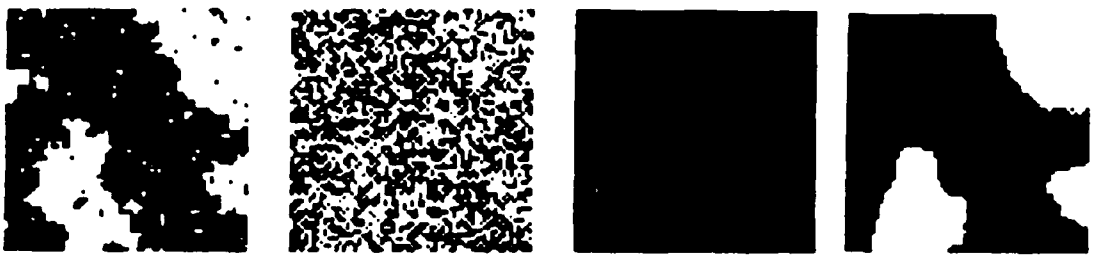


Figure 1. Panel (a) represents a realization of a 64×64 binary Ising net with free boundaries; panel (b), the output of a binary symmetric channel; panel (c), the MAP estimate; and panel (d), an approximation to the MPM estimate.

The main disadvantage of this procedure is that in the case of the segmentation problem, a large amount of memory might be required if the number of classes per element m is large (we need to store the $N(m-1)$ numbers that define the posterior marginals).

With respect to the relative performance, we point out that in many cases, particularly for high signal to noise ratios, the MAP estimate is usually close to the optimal one. If the noise level is high, however, the difference in the performances of the two estimators may be dramatic. This is illustrated in the example portrayed in Figure 1. Panel (a) represents a typical realization of a 64×64 binary Ising net with free boundaries, using a value of $T_0 = 1.74$ (0.75 times the critical temperature of the infinite lattice); panel (b), the output of a binary symmetric channel with error rate $\epsilon = 0.4$; panel (c) the MAP estimate, and panel (d) an approximation to the MPM estimate (which we will label "MPM (M.C.)") obtained using the Metropolis algorithm and Equation (10) to estimate the posterior density. The corresponding values of the posterior energy U_P (Equation (22)) and the relative segmentation error ($e_s/64^2$) are shown on Table 3.

It is clear that the approximation to the MPM estimates shown in panel (d) is better than the MAP from almost any viewpoint.

	f	g	$\hat{f}_{\mathrm{MAP}}$	$\hat{f}_{\mathrm{MPM}}(M.C.)$	$\hat{f}_{\mathrm{MPM}}$ (Det.)
Energy	−5594.8	−226.0	−6660.9	−6460.0	−6247.0
Seg. Error	—	0.4	0.33	0.128	0.124

Table 3.

An intuitive explanation for this behavior comes from the fact that the MAP estimator is implicitly minimizing the expected value of a cost functional $C_{MAP}(f, \hat{f})$ which is equal to zero only if $f_i = \hat{f}_i$ *for all* i, and is equal to, say, M otherwise. If the signal to noise ratio is sufficiently high, the expected value of the optimal segmentation error will be very close to zero, so that $\hat{f}_{MPM}$ and $\hat{f}_{MAP}$ will coincide. In a high noise situation, however, the MAP estimator will tend to be too conservative, since from its viewpoint it is equally costly to make one or one thousand mistakes. The MPM estimator, in contrast, can make a better (although more risky) guess, since making a few mistakes has only a marginal effect on the expected cost.

A quantitative comparison of the performances of the MAP and MPM estimators, with respect to the segmentation error, can be obtained using the ratio:

$$r = \frac{\bar{e}_{MAP}}{\bar{e}_{TPM}} =$$

$$= \frac{\sum_{f,g} \exp\left[-U_P(f;g)\right] e_s(f, \hat{f}_{MAP}(g))}{\sum_{f,g} \exp\left[-U_P(f;g)\right] e_s(f, \hat{f}_{TPM}(g))} \tag{24}$$

In Figure 2 we show a plot of the ratio r for a 2×2 lattice, for different values of the error rate ϵ and the natural temperature T_0. As expected, r is never less than 1. In the worst case (for $\epsilon = 0.1$ and $T_0 = 0.2$) the error of the MAP estimate is 1.17 times that of the MPM estimate: if T_0 is not too small and ϵ is not too large, both estimates coincide, and as ϵ approaches 0.5 (low signal to noise ratio), the MPM estimate is consistently better than the MAP. An experimental analysis of larger lattices reveals a similar qualitative behavior, but the values of r are much larger in this case (see Table 3).

5. Examples of Applications in Vision

5.1. Reconstruction of Piecewise Constant Functions

The efficient solution of this problem is relevant for several reasons: binary im-

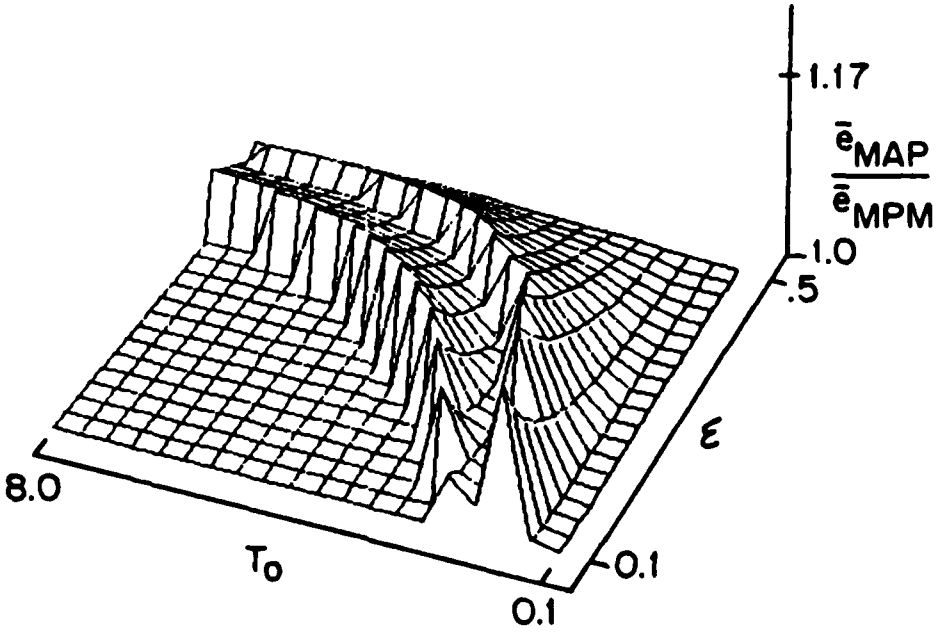


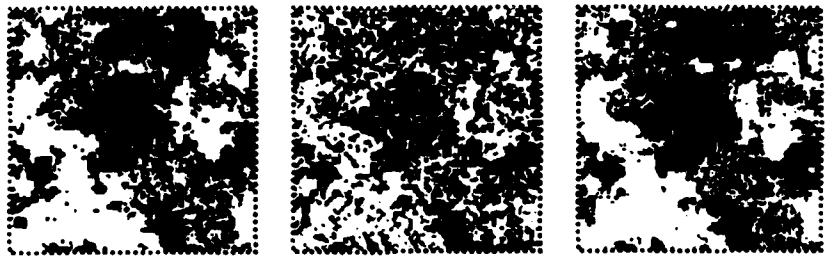
Figure 2. Ratio of the average errors of the MAP and MPM estimators for a 2×2 Ising net.

ages (or images consisting of only a few grey levels) are directly useful in many interesting applications (for example, object recognition and manipulation in restricted (industrial) environments); besides, several perceptual problems, such as the segmentation of textured images (Elliot, et. al. (1983); Hansen and Elliot (1982); Cohen and Cooper (1984)), or the formation of perceptual clusters (Marroquin (1985)) can be reduced to the problem of reconstructing a piecewise constant surface.

The prior model for this kind of functions is given by equations (2) and (6), and the posterior distribution by Equation (8). If the parameters that characterize the system (namely, the "natural temperature" T_0 and the noise parameter α) are known, the MPM estimator produces excellent results, such as the one illustrated in Figure 1.

5.1.1. Parameter estimation

In most practical cases, however, we are only given the noisy observations g and general qualitative information about the structure of the field and the noise, so



(a) (b) (c)

Figure 3. (a) Original ternary MRF. (b) Noisy observations (additive Gaussian noise). (c) Optimal (maximum likelihood) estimate.

that f, the noise parameter α (which is a function of, for example, the error rate ϵ when the noise corruption corresponds to a Binary Symmetric Channel (BSC), or of the variance σ^2, in the case of additive, white Gaussian noise), and the interaction parameter T_0 have to be simultaneously estimated. One plausible approach (Geman, 1985) is to use the "EM algorithm" (Dempster et al., 1977) to find a value for the parameters that maximizes locally the likelihood function:

$$L(\alpha, T_0) = \log P(g|\alpha, T_0) \tag{25}$$

(see Note [1]). For example, for a noise model that corresponds to a BSC with error rate ϵ, the EM algorithm takes the following form. We start with some estimates $\alpha(0), T_0(0)$ for the parameters. The p^{th} iteration (for $p = 1, 2, \ldots$) consists of two steps.

Expectation (E-step): Find the conditional estimates for U_0 and $\hat{\epsilon}$ (see Equations (6) and (9)).

$$\hat{U}_0(p) = E\big[U_0|g, \alpha(p), T_0(p)\big] \tag{26}$$

$$\hat{\epsilon}(p) = E\big[\hat{\epsilon}|g, \alpha(p), T_0(p)\big] \tag{27}$$

These estimates (which are ensemble averages taken with respect to the posterior distribution $P_{f|g}$; see Equation (8)) can be approximated using the Monte Carlo procedure described in Section 4, i.e., computing time averages of the associated ergodic ("Gibbs") chain, generated using the posterior energy U_p.

Maximization (M-step): Find $T_0(p+1), \alpha(p+1)$ such that:

$$E\big[U_0|\alpha(p+1), T_0(p+1)\big] = \hat{U}_0(p); \tag{28}$$

$$E\big[\hat{\epsilon}|\alpha(p+1), T_0(p+1)\big] = \hat{\epsilon}(p). \tag{29}$$

Note that the left hand side of the above expressions – the unconditional expectations of the sufficient statistics U_0 and $\hat{\epsilon}$ – are independent of the data. Thus, the function

$$E[U_0|\alpha, T_0] = E[U_0|T_0] = \Psi(T_0) \tag{30}$$

can be pre-computed for any given lattice size, using again the Monte Carlo procedure of Section 4, but this time with the prior energy U_0 instead of U_p (in Figure 3 we show this function for a 64×64 binary Ising lattice with free boundaries). In this way,

$$T_0^{(p+1)} = \Psi^{-1}(\hat{U}_0^{(p)}) \tag{31}$$

can be computed directly using a table look-up procedure. $\alpha_{(p+1)}$ can be computed directly from the noise statistic $\epsilon_{(p)}$; for example, for a BSC we have:

$$\alpha^{(p+1)} = ln\frac{\hat{\epsilon}^{(p)}}{1-\hat{\epsilon}^{(p)}}. \tag{32}$$

It can be shown that this algorithm will eventually converge towards a fixed point which will be a local maximum of $L(\alpha, T_0)$. Its use, however, has some problems:

Firstly, we note that each iteration of the algorithm – in particular, the E-step – may require a relatively large number of iterations of, say, the Metropolis algorithm. Since the updated values of the parameters $(\alpha(p+1), T_0(p+1))$ are not necessarily close to the old ones $(\alpha(p), T_0(p))$, the Gibbs chain has to be allowed to reach equilibrium each time, before starting to compute the corresponding time averages, which makes the whole procedure computationally expensive. Besides, the likelihood function L will, in general, be multimodal, so that the final result may depend strongly on the choice of the initial values $\alpha(0), T_0(0)$ (in fact, it may be necessary to make several runs with different starting points). Finally, we have found experimentally that the performance of this estimator deteriorates drastically as the SNR falls below a certain level (for example, when the error rate exceeds 0.25, for $T_0 = T_C$ (the critical temperature of the infinite lattice)).

For these reasons, we have used a different approach, which is computationally more efficient, and has better experimental performance. The basic idea is to use some statistics computed from the data to constrain the space of plausible values for the estimates to a smooth curve. In this way we can perform an exhaustive search for the global minimum of an appropriate merit function by varying continuously the values of the parameters, so that the equilibrium of the Gibbs chain is maintained.

To illustrate this idea, we consider the case of a binary Ising field where the noise corruption corresponds to a BSC (the idea can be easily extended to M-ary Ising fields and other noise models - see Note [3]).

We define the statistic U_g as:

$$U_g = \Sigma_{i,j} V(g_i, g_j) \tag{33}$$

where V is an Ising potential (see Section 2). If the error rate of the channel is ϵ, we have that

$$E[U_g|\alpha, T_0] \parallel E[U_0|\alpha, T_o](4\epsilon^2 - 4\epsilon + 1) = \Psi(T_0)(4\epsilon^2 - 4\epsilon + 1) \tag{34}$$

If we make the assumption that

$$E[U_g|\alpha, T_0] = \overline{U}_g \tag{35}$$

where $\overline{U}_g$ is the observed statistic (see Note [2]), we can constrain the search for the estimates $\widehat{\alpha}, \widehat{T}_0$ to the curve given by the equations:

$$\hat{\epsilon} = 1/2[1 - (\overline{U}_g/\Psi(\widehat{T}_0))^{1/2}]$$

$$\hat{\alpha} = ln\Big(\frac{\hat{\epsilon}}{1-\hat{\epsilon}}\Big) \tag{36}$$

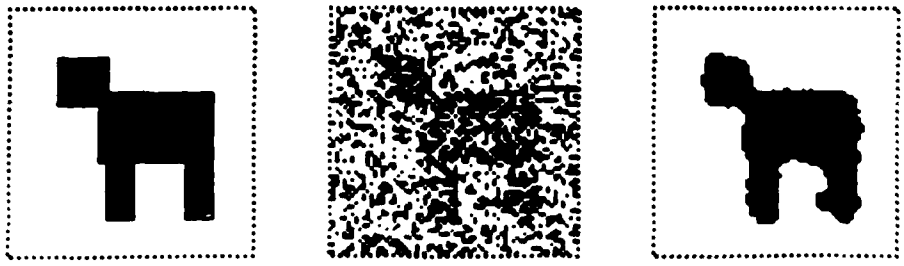
As a merit function we have used the expected degree of uniformity in the spatial distribution of the residuals. In particular, we define a likelihood function L by covering the lattice with a set of m non-overlapping squares (say, 8 pixels wide); computing the relative variance of the noise parameter, estimated over each square, and adding all these terms together:

$$L(\alpha, T_0) = -\sum_{j=1}^{m} \Big(\frac{\hat{\epsilon}_j - \hat{\epsilon}_0}{\hat{\epsilon}_0}\Big)^2, \tag{37}$$

where $\hat{\epsilon}_0$ and $\hat{\epsilon}_j$ are the (conditional) expected values of the noise density over the whole lattice, and over the j^{th} square, respectively, and can be approximated as time averages of the corresponding Gibbs chain (using the posterior energy U_p and α and T_0 as parameters).

The optimal estimate for (α, T_0) can now be obtained as the global minimizer of L over the curve (36). Note that if T_0 (and hence α) are varied slowly enough, so that the associated Gibbs chain is maintained approximately in equilibrium, the computational cost of this search will be equivalent to that of a single "simulated annealing" experiment.

This estimation algorithm allows us to reconstruct a pattern f from the noisy observations g *without having to adjust any free parameters.* The only prior assumptions correspond to the qualitative structure of the field f (first order, isotropic MRF) and to the nature of the noise process. In practice, this means that we can apply it to restore any piecewise uniform image with uniform granularity, even if it has not been generated by a Markov process. We have used



(a) (b) (c)

Figure 4. (a) Synthetic image. (b) Noisy observations. (c) Maximum Likelihood Estimate.

this algorithm to reconstruct a variety of binary images with excellent results. In Figure 4 we show such a restoration. The observations (b) were generated from the synthetic image (a) with an actual error rate of 0.35 (assumed unknown). The optimal estimate for f is shown in (c). It should be noted at this point that for particular problems, it may be possible to find more efficient schemes. Thus, for binary fields sent through a BSC, we have developed a very efficient (deterministic) procedure for approximating the MPM estimator for f which can also be used to find the optimal estimates for α and T_0 (see Marroquin, 1985 for details).

5.2. Reconstruction of Piecewise Continuous Functions.

In this section we will illustrate the application of the local spatial interaction models and estimation techniques that we have described to the reconstruction of piecewise continuous functions from noisy observations taken at sparse locations.

In this reconstruction, it will be important not only to interpolate smooth patches over uniform regions, but to locate and preserve the discontinuities that bound these regions, since very often they are the most important parts of the function. They may represent object boundaries in vision problems (such as image segmentation; depth from stereo; shape from shading; structure from motion, etc.); geological faults in geophysical information processing, etc.

As we mentioned in Section 1.4, an approach to this problem (see Terzopoulos (1984)) consist of first interpolating an everywhere smooth function over the whole domain, then applying some kind of discontinuity detector (followed by a thresholding operation) to try to find the significant boundaries, and finally, reinterpolating smooth patches over the continuous subregions.

The results that have been obtained with this technique, however, are not completely satisfactory. The main problem is that the task of the discontinuity detector is hindered by the previous smooth interpolation operation. This becomes critical when the observations are sparsely located, since in this case, the discontinuities may be smeared in the interpolation phase to such a degree that it may become impossible to recover them in the detection phase.

In contrast, in the Bayesian approach, the boundary detection and interpolation tasks are performed *at the same time*. To apply the general reconstruction algorithms developed above to this problem, the main issue is the representation of the concept of "piecewise continuity" in the form of a prior Gibbs distribution in a meaningful way.

A flexible construction involves the use of two coupled MRF models: one to represent the function (the "surface") itself, and another to model the curves where the field is discontinuous. A coupled model of this kind was first used by Geman and Geman (1984) in the context of the restoration of piecewise constant images. Terzopoulos (1985) has recently attempted to translate this idea in the continuous and deterministic framework of regularization.

This model can be adapted to our problem by modifying the choice of the potentials and the neighborhood structure of the coupled MRF's. Specifically, the following modifications are needed:

1. Since in our case the observations are sparse, it becomes necessary to expand the size of the neighborhoods of the line field, to prevent the formation of "thick" boundaries between the smooth patches (i.e., adjacent, parallel segments of active lines in these regions). In particular, we propose that the dual lattice be 8-connected, with non-zero potentials for the cliques of the form illustrated in Figure 5 (a) and (b). The inclusion of the cliques of Figure 5-b has the additional advantage of penalizing the occurrence of sharp turns, permitting us to model the formation of piecewise smooth boundaries using a binary line process instead of the 4-valued process proposed by Geman and Geman. The potentials for these cliques are computed in the following way:

Let V_a, V_b denote the potentials associated with the cliques C_a, C_b of Figure 5 (a) and (b), respectively, and let S_k ($k \in \{a, b\}$) denote the number of line

Figure 5. Cliques for the line process.

elements belonging to C_k that are "on" at a given time, i.e.,

$$S_k = \sum_{i \in C_k} l_i \quad , \quad k = a, b \tag{38}$$

The potentials V_k are given by:

$$V_k = \beta \phi_k(S_k) \quad , \quad k = a, b \tag{39}$$

where β is a constant, and the functions ϕ_k are defined by the following tables:

S_a	0	1	2	3	4
ϕ_a	0	0.4	0.25	1.2	2.0

S_b	0	1	2
ϕ_b	0	0	10

It is not difficult to see that this choice of potentials will effectively discourage both the formation of thick boundaries ($S_b = 2$) and the presence of sharp turns ($S_a = 3$ and/or $S_b = 2$).

2. The potentials of the depth process, which is now continuous-valued, have to be modified to express the more relaxed condition of piecewise continuity (instead of piecewise constancy). Specifically, we propose:

$$V_(f_i, f_j, l_{ij}) = \begin{cases} (f_i - f_j)^2(1 - l_{ij}), & \text{for } |i - j| = 1 \\ 0, & \text{otherwise.} \end{cases} \tag{40}$$

Note that $l_{ij} \in \{0, 1\}$.

3. Unlike the case of piecewise constant surfaces, we now have to worry about the maximum absolute difference in the values of two adjacent depth sites

that we are willing to consider as a "smooth" gradient (and not a discontinuity). This value, which in general is problem-dependent, determines the magnitude of the constant β in Equation (6), which can be interpreted as the coupling strength between the two processes.

Assuming that the observations are corrupted by i.i.d. Gaussian noise, we get the following expression for the posterior energy:

$$U_P(f,l;g) = \frac{1}{T_0}\sum_{i,j}(f_i - f_j)^2(1-l_{ij}) + \\ + \frac{1}{2\sigma^2}\sum_{i\in S}(f_i - g_i)^2 + \sum_{C_a} V_a(l) + \sum_{C_b} V_b(l), \tag{41}$$

where S is the set of sites where an observation is present. As a performance criterion we will use a mixed cost functional of the form

$$e_m(f,l,\hat{f},\hat{l}) = \sum_{i\in L_f}(f_i - \hat{f}_i)^2 + \sum_{j\in L_l}(1-\delta(l_j - \hat{l}_j)), \tag{42}$$

where L_f, L_l denote the depth and line lattices, respectively. This error criterion means that the reconstructed surface should be as close as possible to the true (unknown) surface, and that we should commit as few errors as possible in the assertions about the presence or absence of discontinuities.

Applying the results of section 3, we find that the optimal estimators will be the *posterior mean* for f and the *maximizer of the posterior marginals* for l.

There is one serious difficulty that prevents us from applying directly the general Monte Carlo procedure that was derived above to the computation of these optimal estimates: since the depth variables are continuous-valued, if we discretize them finely enough to guarantee sufficient precision of the results, the computational complexity of either the Metropolis or Gibbs Sampler algorithms will be very large. One way around this difficulty is to note that for any fixed configuration of the line field, the posterior energy becomes a non-negative definite quadratic form

$$U(f|l,g) = \sum_{i,j:l_{ij}=0}(f_i - f_j)^2 + \alpha\sum_{j\in S}(f_j - g_j)^2 + K \tag{43}$$

where α and K are constants (note that the first sum is taken only over those pairs of sites whose connecting line element is "off," and the second one over the set S). This means that the posterior distribution of the depth field is conditionally Gaussian, so that, for any fixed l, we can find the optimal conditional estimator f_l^* as the mininizer of (25).

Let us define the set F^* as

$$F^* = \{(f,l) \quad : \quad f = f_l^*\}. \tag{44}$$

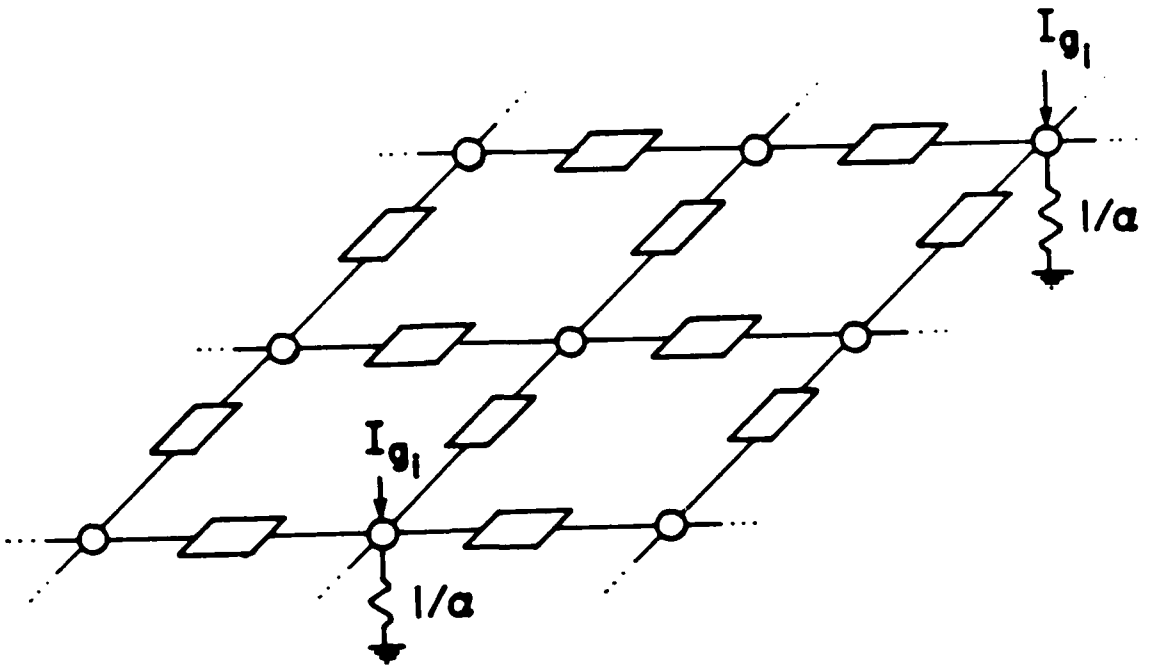


Figure 6. Hybrid network implementing the surface reconstruction algorithm of Section 4. The voltage at every node represents the height of the surface. Inside every rectangular box there is a resistance of unit magnitude and a switch whose state is controlled by the corresponding line element (see text).

It is clear that, if $\hat{f}, \hat{l}$ are the optimal estimates for our problem, we have that

$$(\hat{f}, \hat{l}) \in F^*, \tag{45}$$

which suggests that we can constrain the search for the optimal estimators to this set. This can be done, in principle, by replacing the posterior energy with the function

$$U^*(l) = U(f_l^*, l) \tag{46}$$

(which depends only on l), and use the standard Monte Carlo procedures to find the optimal estimator $\hat{l}$. To illustrate this idea, let us consider a physical model in the next section.

5.2.1. Hybrid Parallel Computers

It is well known that the steady state of an electrical network that contains only (current or voltage) sources and linear resistors will be the global minimizer of a quadratic functional that corresponds to the total power dissipated as heat

(Oster et al, 1971). It is therefore possible to construct an analog network that will find the equilibrium state of the depth field for a given, fixed configuration of the line process, i.e., that will minimize the conditional energy (13) (see Poggio and Koch, 1984; also Poggio et al., 1985). This suggests a hybrid computational scheme in which the line field (whose state is updated digitally, using, say, the Metropolis or Gibbs Sampler algorithms) acts as a set of switches on the connections between the nodes of the analog network whose voltages represent the depth process. In particular, if f_i represents the voltage at node i, the hybrid network can be represented as a 4–connected lattice of nodes (see Figure 6) in which:

A resistance (of unit magnitude) and a switch (controlled by the line element l_{ij}) is present in every link between pairs i, j of adjacent nodes.

If an observation g_i is present at site i, a current of magnitude equal to αg_i is injected to the corresponding node, which must also be connected to a common ground via a resistance of magnitude $1/\alpha$ (see Equation 13).

A direct application of Kirchoff current law shows that at each node i of this network we will have

$$\sum_{j \in N_i} (f_i - f_j)(1 - l_{ij}) + \alpha q_i f_i = \alpha q_i g_i, \tag{47}$$

which corresponds to the condition

$$\text{grad } U(f|l) = 0, \tag{48}$$

so that the equilibrium configuration coincides with f_l^*.

This scheme can be used, in principle, to construct a special purpose hybrid computer for the fast solution of problems of this type. In a digital machine, the exact implementation of this strategy will in general be computationally very expensive, since f_l^* must be computed every time a line site is updated. It is possible, however, to develop approximations which have an excellent experimental performance, and lead to efficient implementations (Marroquin, 1985). The performance of this method is illustrated in Figure 7, in which we show: (with height coded by grey level) the observations (a); the initial state of the network (with all the lines turned "off") (b); the final reconstructed surface (c), and the boundaries found by the algorithm (d), for a square at height 2.0 over a background at constant height = 1.0.

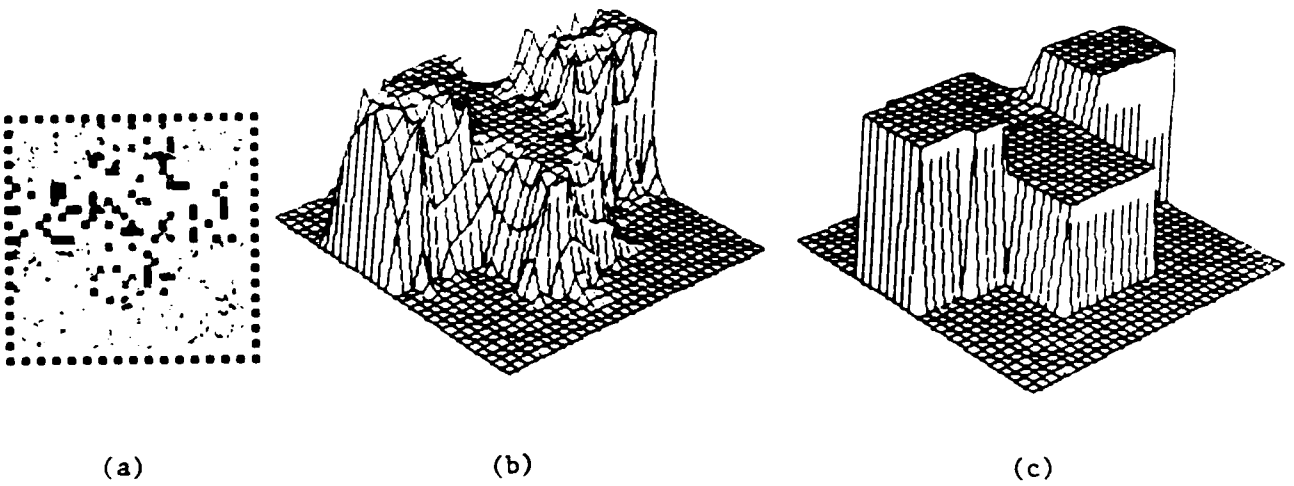


Figure 7. (a) Observations of 3 rectangles at heights 2.0, 3.0 and 2.0 over a background at height 1.0 (height coded by grey level; a white pixel means that the observation is absent at that point). (b) Equilibrium state of the network with all lines turned "off." (c) Optimal estimate.

6. Signal Matching

In all the estimation problems we have studied so far, the posterior energy function had the form

$$U_P(f;g) = U_0(f) + \sum_i \Phi_i(f_i, g_i), \tag{49}$$

where $U_0(f)$ corresponded to the MRF model for the field f. The functions Φ_i, whose precise form depended on the particular noise model, were non–decreasing functions of the distance between f_i and g_i.

There are some cases, however, when the conditional probability distribution of the observations $P_{g|f}(g;f)$ is multimodal (as a function of f) which causes the functions Φ_i to be non–monotonic, so that the solution to the problem remains ambiguous, even if the observations are dense, and the signal to noise ratio arbitrarily high. To illustrate this situation, we will study an important instance of it: the "signal matching" problem, whose one–dimensional version is as follows:

Consider two one–dimensional, real-valued sequences h_L, h_R, where h_L is obtained from h_R by shifting some subintervals according to the "disparity se-

quence" d:

$$h_L(i) = h_R(i + d_i), \tag{50}$$

with

$$d_i \in Q = \{-m, -m+1, \ldots, -1, 0, 1, \ldots, m\}. \tag{51}$$

The signal matching problem is to find d given h_L, h_R. (In a more realistic situation, we do not observe h_L, h_R directly, but rather some noise–corrupted versions g_L, g_R). Some interesting instances of this problem are the matching of stereoscopic images along epipolar lines (Marr and Poggio, 1976); the computation of the dip angle of geological structures from electrical resistivity measurements taken along a bore hole, and the matching of DNA sequences.

To make the discussion more specific, we will consider a simple example, in which the sequences h_L, h_R are binary Bernoulli sequences; we will assume that the noise corruption process can be modeled as a binary symmetric channel with known error rate, and that d is known to be a piecewise constant function. A well known instance of this problem is the matching of a row of a random dot stereogram with density ρ (Julesz (1960)), when the components of the stereo pair are corrupted by noise.

The stochastic model for the observations is then constructed by assuming that the right image is a sample function of a Bernoulli process A with parameter ρ; i.e.,

$$g_R(i) = A(i). \tag{52}$$

The left image is assumed to be formed from the right one by shifting it by a variable amount given by the disparity function d, except at some points where an error is committed with probability ϵ. Note that some regions that appear in the right image will be occluded in the left one (see Figure 8). The "occlusion indicator" ϕ_d can be computed deterministically from d in the following way:

$$\phi_d(i) = \begin{cases} 1, & \text{if } d_{i-k} \geq d_i + k, \text{ for some integer } k \in (0, m) \\ 0, & \text{otherwise.} \end{cases} \tag{53}$$

The occluded areas are assumed to be "filled in" by an independent Bernoulli process B. The final model is then:

$$g_L(i) = \begin{cases} g_R(i + d_i) & \text{with prob. } 1 - \epsilon, \text{ if } \phi_d(i) = 0 \\ 1 - g_R(i + d_i) & \text{with prob. } \epsilon, \text{ if } \phi_d(i) = 0 \\ B_\rho(i) & \text{with prob. 1, if } \phi_d(i) = 1. \end{cases} \tag{54}$$

Note that in the two dimensional case, the index i denotes a site of a lattice, and therefore it can be represented as a two–vector (i_1, i_2) whose components denote the column and row of the site, respectively. To simplify the notation, we will adopt the following convention throughout this section: when a scalar is added to this vector index (as in $g_R(i + d_i)$ and d_{i+1}), it will be implicitly

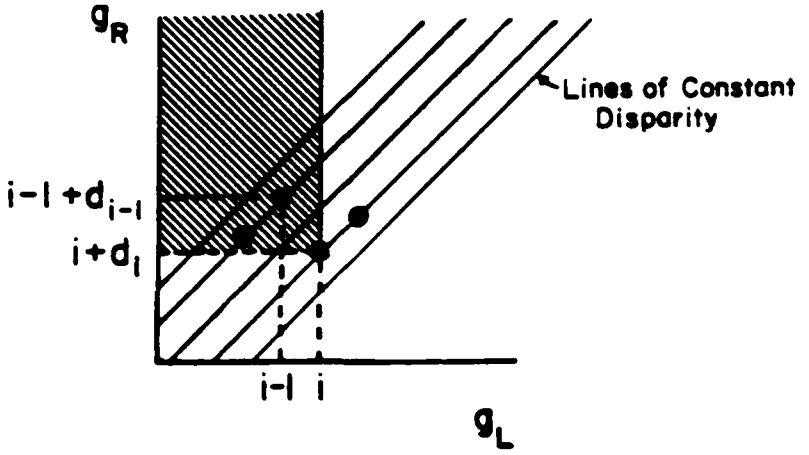


Figure 8. Occluded Regions: The horizontal and vertical axis represent points in one row of the left and right images, respectively. Matching points are represented by black circles. Any match in the shaded region will occlude the point i.

assumed that it is multiplied by the vector $(1,0)$ (so that the above expressions should be understood as $g_R(i + (d_i, 0))$ and $d_{i+(k,0)}$, respectively). Using this convention, the observation model of Equation (27) can be applied either to the one or to the two–dimensional cases.

Notice that even if the observations are noise–free ($\epsilon = 0$) the solution of the problem remains ambiguous, and it cannot be uniquely determined unless some prior knowledge about d (for example, in the form of a MRF model) is introduced. The use of a MRF model in the stereo matching case corresponds to a quantification of the assumption of the existence of "dense solutions" (this term was introduced by Julesz (1960), and essentially corresponds to the assumption that the disparity d varies smoothly in most parts of the image; see also Marr and Poggio (1979)), and the use of the occlusion indicator corresponds to the "ordering constraint" (i.e., the requirement that if $i > j$, then $i + d_i > j + d_j$, see Baker (1981); we put $\phi_d = 1$ whenever this constraint is violated).

To formulate the estimation problem, we will consider the sequence g_L as "observations," while g_R will play the role of a set of parameters. Thus, from (27), we have (assuming, for simplicity that $\rho = \frac{1}{2}$):

$$P(g_L(i) = k|d, g_R) = P_{g|d}(k) = \begin{cases} 1-\epsilon & \text{if } \phi_d(i) = 0 \text{ and } g_R(i+d_i) = k \\ \epsilon & \text{if } \phi_d(i) = 0 \text{ and } g_R(i+d_i) \neq k \\ \frac{1}{2} & \text{if } \phi_d(i) = 1 \end{cases} . \tag{55}$$

As a prior model for the disparity field, we may use a first order MRF with generalized Ising potentials, such as the one presented in Section 5.1. Other models may also be used, including the coupled depth and line fields that we discussed in the previous section. For the present, let us assume that the simpler Ising model is adequate. Note that even when the matching problem is one-dimensional (we are assuming that there is no vertical disparity between the images, so that the matching can be done on a row-by-row basis), the two-dimensional nature of the prior MRF model for the disparity introduces a coupling between matches at adjacent rows. The posterior energy is

$$U_P(d;g) = \frac{1}{T_0}\sum_{i,j} V(d_i,d_j) + \frac{1}{2}\sum_i \phi_d(i)\ln 2 + \\ + \frac{\alpha}{2}\sum_i (1-\phi_d(i))\delta(g_L(i) - g_R(i+d_i)), \tag{56}$$

where

$$\alpha = \ln\left(\frac{\epsilon}{1-\epsilon}\right). \tag{57}$$

It is possible to apply the general Monte Carlo algorithms presented above to approximate the optimal estimate $\hat{d}$ with respect to a given performance measure (such as the mean squared error). Their use in this case, however, is complicated by the introduction of the occlusion function ϕ_d in the posterior energy: the size of the support for this function equals the total number of allowed values for the disparity (see Equation (36)). If this number is large, the computation of the increment in energy, or of the conditional distributions (if the Gibbs Sampler is used) may be quite expensive. In many cases, however, the size of the regions of constant disparity is relatively large compared with the size of the occluded areas. In these cases, one can approximate the posterior energy by:

$$U_P(d) = \frac{1}{T_0}\sum_{i,j} V(d_i,d_j) + \frac{\alpha}{2}\sum_i \delta(g_L(i) - g_R(i+d_i)) \tag{58}$$

and increase significantly the computational efficiency. Notice that this approximate functional is very similar to the functional suggested by standard regularization (compare Table 2). It is also possible, particularly for the high signal to noise ratio case, to design deterministic, highly distributed algorithms for the efficient computation of the optimal estimator. The details of these designs can be found in Marroquin, 1985.

To illustrate the performance of this approach, we present in Figure 9 a random dot stereogram portraying a square floating over a uniform background (panel (a)), and the reconstructed surface (panel (b)).

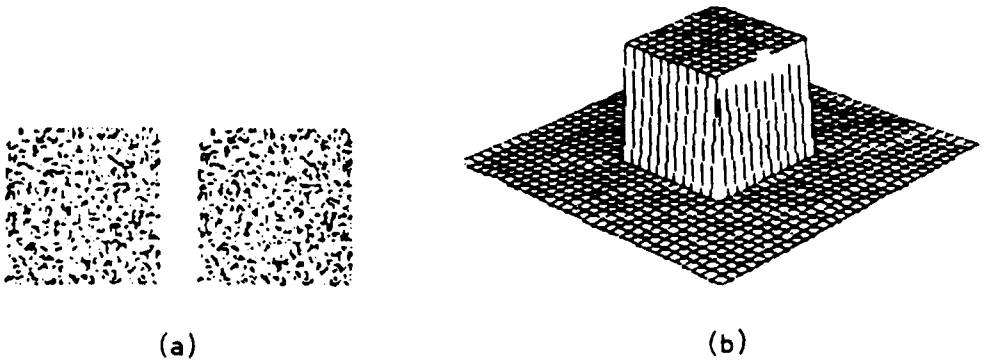


Figure 9. (a) Random dot stereogram. (b) Reconstructed surface.

7. Parallel Implementations.

7.1. Connection Machine Architectures.

The general Monte Carlo procedure that we have presented for the approximation of the optimal Bayesian estimators of MRF's can be greatly accelerated if it is implemented in a parallel architecture. A necessary condition for the convergence of the probability measures of the Markov chains defined by the Metropolis, Gibbs Sampler, or heat bath algorithms to the posterior Gibbs distribution (and therefore, for the convergence of the approximations given by Equations (22) and (23) to the desired estimates) is that if two sites belong to the same clique, they are never updated at the same time. It is important to note, however, that this condition is also sufficient only for the case of the Gibbs sampler and heat bath algorithms: if one updates simultaneously the states of all non-neighboring sites, the regularity of the resulting Metropolis chain will be destroyed, so that it will no longer be possible to guarantee the convergence of the Metropolis algorithm to the desired result (see Note [4]).

If one implements the Gibbs sampler in a parallel architecture in which a processor is assigned to each site, the total execution time will be reduced by a factor of

$$\frac{N}{K} \tag{59}$$

where K is the so called "chromatic number" of the graph that describes the neighborhood structure, and it is equal to the minimum number of colors needed to color the sites of the lattice in such a way that no two neighbors are the same.

An example of such a massively parallel architecture is the "Connection Machine" computer (Hillis, 1985), built by Thinking Machines Corporation.

This machine was originally conceived at the Artificial Intelligence laboratory at MIT and was further developed and now marketed by Thinking Machines Corporation. It is a "Single Instruction Multiple Data" (SIMD) array processor consisting in the version presently available at the AI Laboratory of 64,000 processing units (each with a single bit Arithmetic/Logical unit) organized in a four–connected lattice that is 128 elements square. Besides this nearest–neighbor connectivity, it will also be possible (although computationally more expensive), to connect any two processors in the array using a router network.

At each cycle of the machine an instruction is executed by each processor, and a single bit is transmitted to its neighbors. This means that the updating scheme can be implemented most efficiently if the field is first order Markov, but higher order processes can also be implemented without using the router by successively propagating the transmitted state (the execution time, therefore, will grow linearly with the order of the field).

To make this discussion more concrete, consider, as an example, the problem of finding the optimal estimate for an M–ary, first order MRF with Ising potentials (i.e., the segmentation of a piecewise constant image) from noisy observations. Let us assume that the estimator is to be implemented in the Connection Machine system, and suppose that by the use of appropriate scaling factors, all the numbers can be represented as 16–bit integers. We will use the following conservative assumptions: We assume that 16 cycles of a single 1–bit processor are needed to perform 16–bit addition, subtraction or comparison; 16^2 cycles to perform multiplication or division; 2×16^2 cycles for generating a pseudo–random number with uniform distribution on a given interval; 16 cycles for memory transfer operations, and 6×16^2 cycles for computing an exponential.

Assuming that we run 250 iterations of the system, and ignoring the overhead time we get that

$$\text{Exec. Time} \approx 1.4(M-1)10^6 \text{ cycles} \tag{60}$$

For the particular case of binary images, we have developed a deterministic scheme for which this execution time can be reduced by an order of magnitude (see Marroquin, 1985).

In the case of the reconstruction of piecewise smooth functions from sparse data, the optimal estimator can also be implemented in this machine. To study this implementation, we first note that the chromatic numbers of the graphs associated with the line and depth neighborhood systems are 4 and 2, respectively,

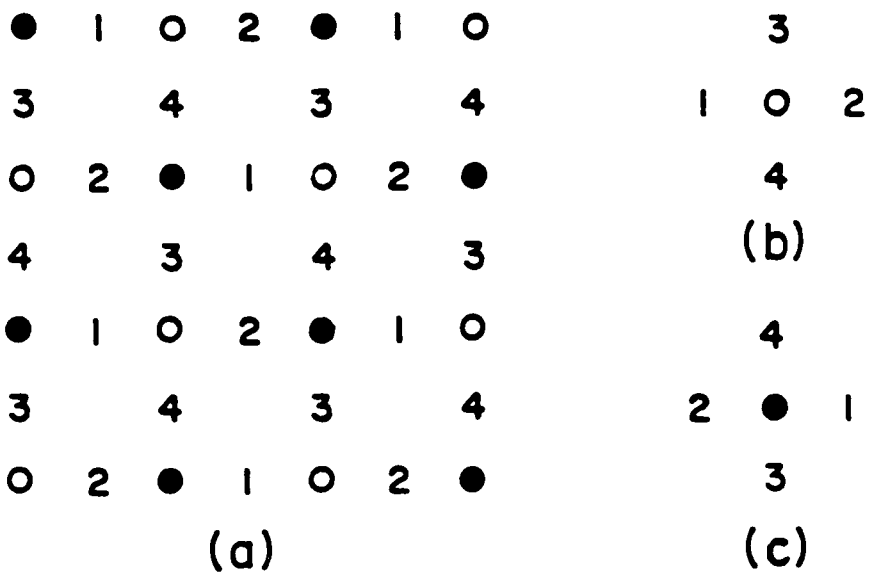


Figure 10. (a) Coloring of the coupled line–depth lattice. (b) and (c) Elements whose state is stored in each of the two types of processors of a 4–connected parallel architecture.

which means that the coupled process has a chromatic number of 6. In Figure 10 (a) we illustrate one possible "coloring."

The colors of the line process are represented by the numbers 1,2,3,4, and those of the depth process by white and black circles. The updating process can be implemented in a 4–connected architecture by assigning one processor to each depth site and its four adjacent line elements. We will thus have two different populations of processors, whose configurations are shown in Figures 9 (b) and (c), respectively.

Each complete iteration consist on 6 major cycles: in the first two, the state of the white and black depth variables is respectively updated, and in the next four, the new states of the binary line variables stored in (say) the white processors are successively computed and transmitted to the corresponding memory locations of the neighboring black processors. Note that in this scheme we have some redundancy in the use of memory (each binary variable is stored twice), but the state of all the elements needed for each updating operation is always available from adjacent processors. The Monte Carlo algorithm requires about 200 iterations to converge. As before, we have also developed in this case a deterministic scheme with very good experimental performance, for which the

execution time can be reduced by at least an order of magnitude. E. Gamble has now implemented several MRF models on the Connection Machine system at the AI Laboratory as part of the Vision Machine project (Poggio et al., 1987).

7.2. Hybrid Analog-Digital Computers and Hopfield Networks

As we mentioned in section 4.2.1, the reconstruction of piecewise continuous functions can be achieved by coupling two MRF's, one corresponding to the continuous field and the other to the discontinuities. From this scheme we have suggested a special purpose parallel computer consisting of an analog network of resistances – corresponding to the continuous intensity field – and a digital network – corresponding to the line process, coupled *via* D-A and A-D converters. The idea suggested by computer experiments (Marroquin, 1985) is that the two processes can run on different time scales, a slow one for the digital part and a fast one for the analog network. In this way the two processes are effectively decoupled and the continuous field finds its equilibrium effectively instantaneously after each update of the line process. (Koch, Marroquin and Yuille (1985) discuss implementations of this idea.) It can be extended to ***multilayered hybrid networks***, each layer corresponding to a MRF and being digital or analog depending on the continuous or binary nature of the field. Hybrid multilayered architectures of this type are especially attractive for implementing the fusion of several vision processes.

Finally, we mention that Koch, Marroquin and Yuille (1985) have been experimenting successfully with a special type of analog networks - Hopfield networks – whose equilibrium states correspond to approximations of the optimal estimators.

8. Conclusions

In this paper we have presented a probabilistic approach to the solution of a class of perceptual problems. We showed that these problems can be reduced to the reconstruction of a function on a finite lattice from a set of degraded observations, and derived the Bayesian estimators that provide an optimal solution. We have also developed efficient distributed algorithms for the computation of these estimates, and discussed their implementation in different kinds of hardware. To demonstrate the generality and practical value of this approach, we

studied in detail several applications: the segmentation of noise–corrupted images; the reconstruction of piecewise smooth surfaces from sparse data and the reconstruction of depth from stereoscopic measurements.

8.1. Connection with Standard Regularization

The maximum *a posteriori* (MAP) estimate of a MRF is obviously similar to a variational principle of the general form of Equation (7), since the use of this criterion defines the optimal estimator as the global minimizer of the posterior energy U_P (Equation (11)): the first term measures the discrepancy between the data and the solution, the second term is now an arbitrary "potential" function of the solution (defined on a discrete lattice). It is then natural to ask for the connection between standard regularization principles and the MRF approach. It turns out (from Equation (8)) that a MAP estimate leads to the minimization of a functional U_P (see Equation (9))– in general not quadratic – that *reduces to a quadratic functional, of the standard regularization type, when the MRF is continuous-valued, the noise is additive and gaussian* (the term $\sum \Phi_i(f, g_i)$ will be quadratic) *and first order differences of the field are zero-mean, independent, gaussian random variables* (thus the a priori probability distribution is a Gibbs distribution with quadratic potentials so that the term the term $U_0(f)$ is quadratic).

8.2. The Fusion Problem

This approach also permits, in principle, the incorporation of more than one modality of observations into a single estimation process, as well as the simultaneous estimation of several related functions from the same data set. This makes one hope that this framework could be useful in the solution of difficult problems that require such an integrated approach.

For instance, the stereo matching problem in real situations has not been solved yet in a completely satisfactory way. The same can be said of other related perceptual problems such as: edge detection; image segmentation; the recovery of the shape of an object from a single two–dimensional image (the "shape form shading" problem), and the segmentation of a scene into distinct objects, as well as the recovery of their three–dimensional structure from the analysis of images formed at successive instants of time (the "structure from motion" problem). All these problems are obviously related, and it is intuitively clear that the individual solutions that can be obtained should improve if the mutual constraints that the solution of each individual problem imposes on the

others were taken into account. Thus, the presence of a brightness edge should increase the likelihood of a depth edge, and vice versa; the depth estimated from stereo should be compatible with the shape derived from shading; points belonging to the same region in an image should move together, etc. We believe that these constraints can be incorporated in the potential functions of the corresponding MRF models, so that the combined optimal estimation process represents, in fact, an integrated cooperative solution to these problems, with a significantly improved performance. Recently, Poggio has discussed how to use coupled MRF models to integrate information from different vision modules (Poggio, 1985). Gamble and Poggio (in Poggio et al., 1987) have implemented efficient algorithms to integrate stereo, motion and intensity information to obtain a robust description of the discontinuities in the scene.

9. Notes

[1]: It is computationally unfeasible to perform the maximization of the likelihood function L directly, due to the extraordinary complexity of $P(g|\alpha, T_0)$:

$$P(g|\alpha, T_0) = \frac{\sum_f \exp\left[-U_p(f; g|\alpha, T_0)\right]}{\sum_{f,h} \exp\left[-U_p(f; h|\alpha, T_0)\right]} \tag{61}$$

where U_p is given by Equation (9). However, the form of the "complete data" distribution $P(f, g|\alpha, T_0)$ (the so-called "regular exponential family form"; see Dempster et al., 1977) is such that at every local maximum of L we have that:

$$\begin{aligned} E[U_0|g, \alpha, T_0] &= E[U_0|\alpha, T_0] \\ E[\hat{\epsilon}|g, \alpha, T_0] &= E[\hat{\epsilon}|\alpha, T_0] \end{aligned} \tag{62}$$

where $\hat{\epsilon}$ and U_0 are defined by Equations (6) and (7).

Note that both the left and the right hand sides of the above equations can be approximated using the Monte Carlo procedure described in Section 4 (using the posterior and prior energy, respctively), and that the right hand side is independent of the observations. These relations form the basis of the EM algorithm.

[2]: Since both the random field f and the noise process are stationary, we have that

$$E\left[\left(\hat{U}_g - E[\hat{U}_g|\alpha, T_0]\right)^2\right] = \frac{1}{\#\text{ of cliques of the lattice}} \tag{63}$$

so that this assumption becomes asymptotically correct for large lattices.

[3]: Consider an M-ary field f with Ising potentials, corrupted with O-mean, additive white Gaussian noise with variance $\sigma^2 < \sigma^2_{\max}$. Suppose that

$$f_j \epsilon Q = \{q : q = q_0 + 2k\partial, k = 1, 2, \ldots, M\} \text{ for all } i. \tag{64}$$

We define the statistic W_g as

$$W_g = \frac{1}{N_C} \sum_{i,j} W(G_i, g_j), \tag{65}$$

where g is the observation process, N_C is the number of nearest-neighbor pairs in the lattice,

$$W(g_i, g_j) = \begin{cases} -1, & \text{if } \hat{g}_i = \hat{g}_j \text{ and } |i-j| = 1 \\ 1, & \text{if } \hat{g}_i \neq \hat{g}_j \text{ and } |i-j| = 1 \\ 0, & \text{if } |i-j| \neq 1, \end{cases} \tag{66}$$

and $\hat{g}i = q_0 + 2n\partial$, with n an integer such that

$$q_0 + (2n-1)\partial < g_i \leq 0 + (2n+1)\partial. \tag{67}$$

Note that it is possible that $\hat{g}_i \notin Q$.

Define

$$\varphi(r, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \int_{\infty}^{r} \exp[-x^2/2\sigma^2] dx. \tag{68}$$

It is not difficult to see that

$$E[W_g|\sigma, T_0] = 1 - (A + B) + E[U_0|T_0](A - B) \tag{69}$$

where $\overline{U}_0 = U_0/N_C$;

$$\begin{aligned} A &= Pr(W_g(g_i, g_j) = -1 | V(f_i, f_j) = -1), \\ B &= Pr(W_g(g_i, g_j) = -1 | V(f_i, f_j) = 1) \end{aligned} \tag{70}$$

(note that $E[U_0|T_0] = \Psi(T_0)$ is data independent, and therefore, it can be computed off-line).

Assuming that

$$\frac{Pr(|f_i - f_j| = q | f_i \neq f_j)}{M-1}, \quad \text{for } q = 1, 2, \ldots, M-1, \tag{71}$$

we can approximate A and B by:

$$\begin{aligned} A(\sigma) &= 6a^2 + 4b^2 - 4ab - 4a + 1 \\ B(\sigma) &= \frac{1}{M-1}(-3a^2 - 3b^2 + 2ab + 2a), \end{aligned} \tag{72}$$

where $a = \varphi(\partial, \sigma)$ and $b = \phi(3\partial, \sigma)$. (The above approximation has been computed assuming that $\phi(5\partial, \sigma_{\max}) \approx 0$. If this is not true, more terms can easily be included.)

Assuming, as before, that

$$E[W_g|\sigma, T_0] = \overline{W}_g \quad \text{(computed from the data)}, \tag{73}$$

we can find the optimal estimate for (σ, T_0) as the global maximizer of an appropriate merit function along the curve

$$T_0 = \frac{\Psi^{-1}\overline{W}_g + A(\sigma) + B(\sigma) - 1}{A(\sigma) - B(\sigma)}, \tag{74}$$

using a "composite annealing" strategy.

To define the merit function, we cover the lattice with a set of non-overlapping squares and add the relative variance of the noise parameter over each square, thus:

$$\angle(\sigma, T_0) = -\sum_{j=1}^{m} \left(\frac{\sigma_0 - \sigma_j}{\sigma_0}\right)^2. \tag{75}$$

σ_0 and σ_j can be approximated as time averages, with respect to the Gibbs chain with σ and T_0 as parameters, of the estimated noise variance over the whole lattice and over the jth square, respectively.

[4]: As a simple example for which the regularity of the Metropolis chain is destroyed, consider a 3×3 ginary Ising lattice with periodic boundary conditions. It is easy to see that for the initial state:

$$\begin{array}{ccc} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{array}. \tag{76}$$

The Metropolis algorithm, either with lexicographic updating order, or with simultaneous updating of all non-neighboring sites, will produce, deterministically, the sequence:

$$\begin{array}{ccccccccccccc} 1 & 0 & 1 & & 0 & 1 & 0 & & 1 & 0 & 1 & & \\ 1 & 0 & 1 & \rightarrow & 0 & 1 & 0 & \rightarrow & 0 & 1 & 0 & \rightarrow & \ldots \\ 0 & 1 & 0 & & 1 & 0 & 1 & & 1 & 0 & 1 & & \end{array} \tag{77}$$

for any finite temperature.

Acknowledgements

We are grateful to Eric Grimson and Jim Little for reading the paper and making several useful suggestions.

Reading list

Abend, K. "Compound decision procedures for unknown distributions and for dependent states of nature," in **Pattern Recognition**, L. Kanal, ed., Washington, D.C.: Thompson Book Co. (1968).

Barrow, H.G. and J.M. Tennenbaum. "Interpreting line drawings as three dimensional surfaces," *Artificial Intelligence*, **17**, 1981.

Bertero, M., Torre, V. and Poggio, T. "Ill-posed problems in early vision" A.I. Laboratory Memo 924, 1986.

Besag, J. "Spatial interaction and the statistical analysis of lattice systems," *J. Royal Stat. Soc. B*, **34**, 75-83 (1972).

Cohen, F.S., and D.B. Cooper. "Simple parallel hierarchical and relaxation algorithms for segmenting noncausal Markovian random fields," Brown University Laboratory for Engineering Man/Machine Systems, Tech. Report LEMS-7 (1984).

Cross, G.C. and A. K. Jain. "Markov random field texture models," *IEEE Trans. Pattern Analysis and Machine Intelligence*, **5** (1983).

Dempster, A.P., N.M. Laird and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Stat. Soc. B*, **39**, 1-38 (1977).

Elliot, H., R. Derin, R. Christi, and D. Geman, "Application of the Gibbs distribution to image segmentation," Univ. of Massachusetts Technical Report (1983).

Feller, W. **An Introduction to Probability Theory and its Applications**, Vol I., New York: John Wiley and Sons (1950).

Gallager, R.G. **Information Theory and Reliable Communication**, New York: John Wiley and Sons (1968).

Geman, S. and D. Geman. "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, **6**, (1984).

Geman, S. Personal communication (1985).

Grenander, U. **Tutorial in Pattern Theory**, Div. of Applied Math., Brown University (1984).

Grimson, W.E.L. **From Images to Surfaces**, Cambridge, MA: M.I.T. Press (1981).

Grimson, W.E.L. "A computational theory of visual surface interpolation," *Phil. Trans. R. Soc. London B*, **298**, (1982a).

Habibi, A. "Two dimensional Bayesian estimation of images," *Proc. IEEE*, **60**, (1972).

Hansen, A.R. and H. Elliot. "Image segmentation using simple Markov field models," *Comp. Vision, Graphics, and Image Proc.*, **20**, (1982).

Hassner, M. and J. Sklansky. "The use of Markov random fields as models of texture," *Comp. Vision, Graphics and Image Proc.*, **12**, (1980).

Hillis, D. "The Connection Machine," Ph.D. Thesis, M.I.T. Department of Electrical Engineering and Computer Science (1985).

Horn, B.K.P. *Robot Vision*, Cambridge and New York: MIT Press and McGraw-Hill, 1986.

Julesz, B. "Binocular depth perception of computer generated patterns," *Bell Sys. Tech. J.*, **39** (1960).

Kashyap, R.L., and R. Chellappa. "Estimation and choice of neighbors in spatial interaction models of images," *IEEE Trans. on Info. Theory*, **29** (1983).

Kemeny, J.G., and J.L. Snell. **Finite Markov Chains**, New York: Van Nostrand (1960).

Kirkpatrick, S., C.D. Gelatt, and M.P. Vecchi, "Optimization by simulated annealing," *Science*, **220** (1983).

Koch, C., J. Marroquin, and A. Yuille. "Analog 'neuronal' networks in early vision," Massachusetts Institute of Technology Artificial Intelligence Laboratory Memo 751 (1985).

Marr, D. **Vision: A Computational Investigation into the Human Representation and Processing of Visual Information**, San Francisco: W.H. Freeman and Co. 1982.

Marr, D. and T. Poggio. "From understanding computation to understanding neural circuitry," *Neur. Res. Bull.*, **15** (1977).

Marroquin, J. "Surface reconstruction preserving discontinuities," Massachusetts Institute of Technology Artificial Intelligence Laboratory Memo 792 (1984).

Marroquin, J. "Probabilistic solution of inverse problems," Ph.D. Thesis, Massachusetts Institute of Technology (1985).

Metropolis, N. et al. "Equation of state calculations by fast computing machines," *J. Phys. Chem.*, **21** (1953).

Morozov, V.A. **Methods for Solving Incorrectly Posed Problems**, New York: Springer-Verlag 1984.

Nahi, N.E. and T. Assefi. "Bayesian recursive image estimation," *IEEE Trans. on Computers*, **21** (1972).

Oster, G.F., A. Perelson and A. Katchalsky. "Network *thermodynamics*," *Nature*, **234** (1971).

Poggio, T. "Vision by man and machine," Massachusetts Institute of Technology Artificial Intelligence Laboratory Memo 776 (1984).

Poggio, T. and C. Koch. "Analog networks: a new approach to neural computation," Massachusetts Institute of Technology Artificial Intelligence Laboratory Memo 783. (1984).

Poggio, T. "Integrating Vision Modules with Coupled MRF's," Massachusetts Institute of Technology Artificial Intelligence Laboratory Working Paper 285, 1985.

Poggio, T. and V. Torre. "Ill-posed problems and regularization analysis in early vision," Massachusetts Institute of Technology Artificial Intelligence Laboratory Memo 773 (1984).

Poggio, T., V. Torre, and C. Koch. "Computational vision and regularization theory," *Nature*, **317** (1985).

Poggio, T., H. Voorhees, and A. Yuille. "Regularizing edge detection," Massachusetts Institute of Technology Artificial Intelligence Laboratory Memo 776 (1984).

Poggio, T. and staff " MIT progress in understanding images,". In: Proceedings of the Image Understanding Workshop, L. Bauman (ed.), Science Applications International Corp., McLean, VA, 1987.

Terzopoulos, D. "Multiresolution computation of visible-surface representations," Ph.D. Thesis, M.I.T. Department of Electrical Engineering and Computer Science (1984).

Terzopoulos, D. "Integrating visual information for multiple sources for the cooperative computation of surface shape," in **From Pixels to Predicates: Recent Advances in Computational and Robotic Vision**, A. Pentland, ed. Ablex (1986).

Tikhonov, A.N., and V. Y. Arsenin. **Solutions of Ill–Posed Problems**, New York: Winston and Sons (1977).

Winston, P. "Proposal to DARPA," Massachusetts Institute of Technology (1984).

Wong, E. "Two-dimensional random fields and the representation of images," *SIAM J. App. Math.*, **16**, 4 (1968).

Woods, J. W. "Two-dimensional discrete Markovian fields," *IEEE Trans. Info. Theory*, **18** (1972).

END

9-87

DTIC